

Capstone Project 1 - EDA

Hotel Booking Analysis

Supriya Nammi

The Big Question

Hotel Managements the world over, strive to have the maximum number of bookings at the highest possible price.

Customers, on the other hand, would like to book the Hotel and rooms of their choice, at the lowest possible price.

The question is- Can data help?

Exploratory Data Analysis (EDA) of Hotel booking data can provide important insights to help both, the Hotel Management as well as the Customers, get what they want.

So, let us see what data do we have, and how we can help.

The Data Dictionary

We have a dataset that contains 119,390 observations and 32 features

Numerical Data

- lead_time
- stays_in_weekend_nights
- stays_in_week_nights
- adults
- children
- babies
- previous_cancellations
- previous_bookings_not_canceled
- booking_changes
- days_in_waiting_list
- adr
- required_car_parking_spaces
- total_of_special_requests

Categorical Data

- hotel
- is_canceled
- arrival_date_year
- arrival_date_month
- arrival_date_year_week_number
- arrival_date_day_of_month
- market_segment
- distribution_channel
- reservation_status
- reservation_status_date
- is_repeated_guest
- meal
- country
- deposit_type
- agent
- company
- customer_type
- reserved_room_type
- assigned_room_type

Unique Values in Categorical Columns

- **hotel** : Resort Hotel, City Hotel
- **meal** : BB (Bed & Breakfast), FB (breakfast, lunch and dinner), HB (breakfast and one other meal), SC (no meal package), Undefined
- **market_segment** : Direct, Corporate, Online TA, Offline TA/TO (Travel Agency/ Tour Operator), Complementary, Groups, Aviation, Undefined
- **distribution_channel** : Direct, Corporate, TA/TO, Undefined, GDS
- **reserved_room_type** : C, A, D, E, G, F, H, L, P, B
- **assigned_room_type** : C, A, D, E, G, F, I, B, H, P, L, K
- **deposit_type** : No Deposit, Refundable, Non Refund
- **customer_type** : Transient, Contract, Transient-Party, Group
- **reservation_status** : Check-Out, Canceled, No-Show

What can we learn from the available data?

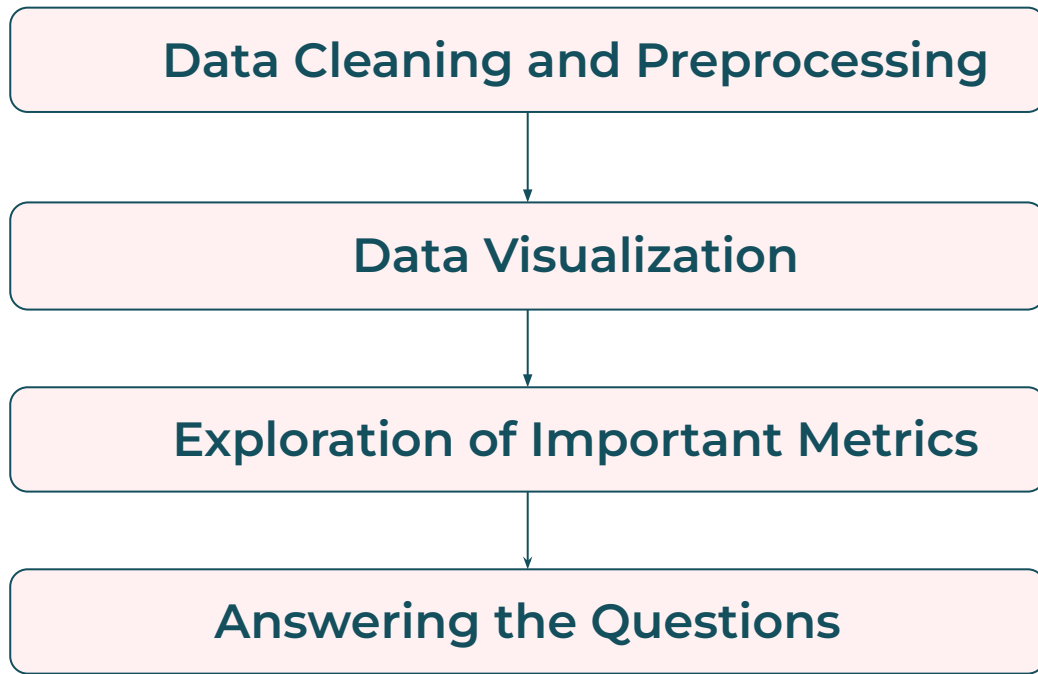
Some important questions from the customer's perspective that can be answered are:

- When is the best time of the year to book the hotel rooms?
- Would room be assigned according to choice?
- What is the optimal length of stay in order to get the best daily rate?

Questions from the Hotels' perspective:

- When is the hotel more likely to receive high number of special requests?
- How can the ADR be increased?
- What are the reasons behind booking cancellations ?

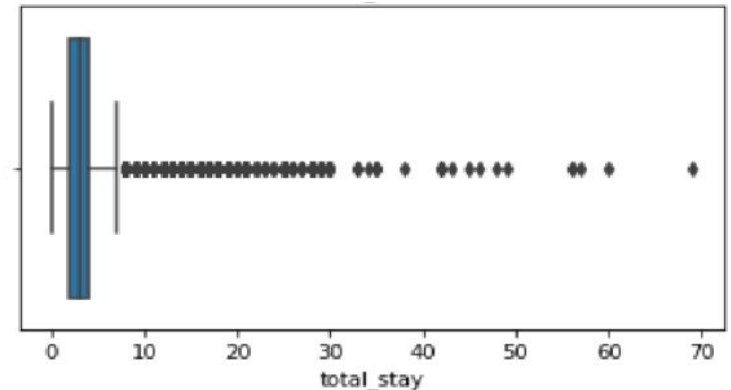
Trajectory of the EDA



Data Cleaning and Preprocessing

- Removal of missing/null values.
- Removal of unwanted columns/features and Creation of new ones.
- Handling of outliers: Imputation of a maximum value for all the extreme points in the columns.

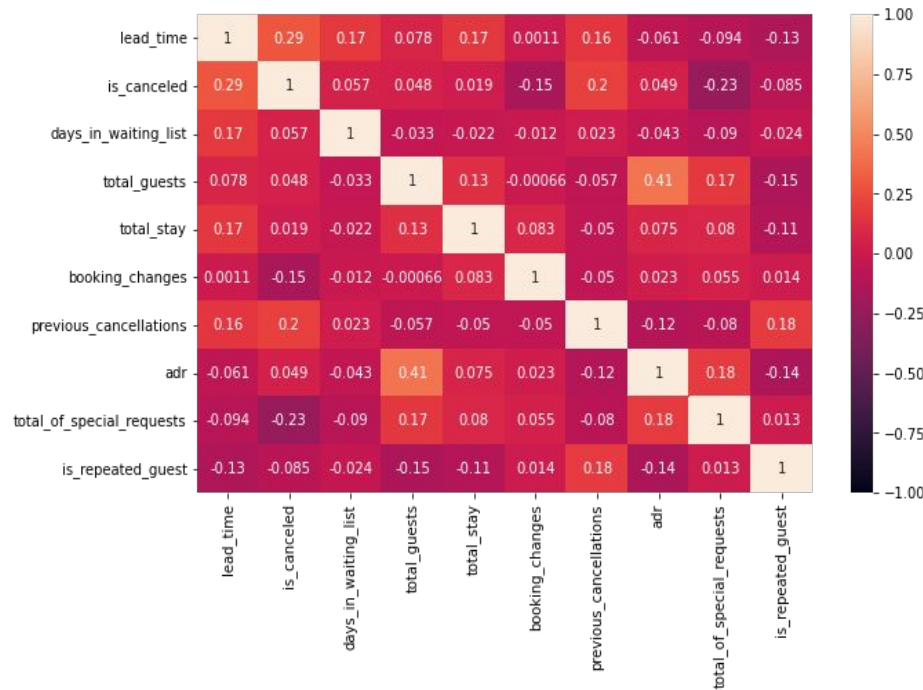
	Missing Values	% Missing
children	4	0.00
country	488	0.41
agent	16340	13.69
company	112593	94.31



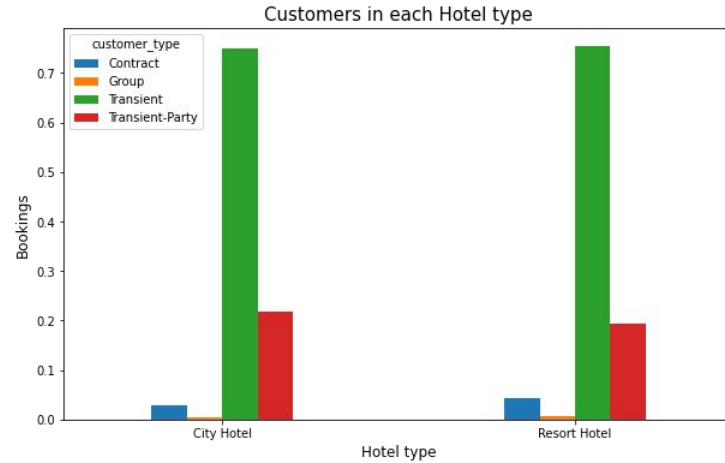
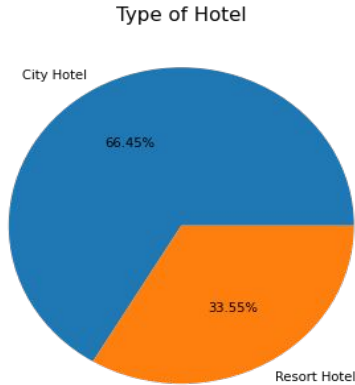
Data Visualization

Correlation Heat Map

- 'adr' (Average Daily Rate) has the highest correlation (0.41) with 'total_guests'.
- Lead time seems to have some bearing on cancellations.
- Rest of the features in the dataset do not contain very strong correlation between each other.

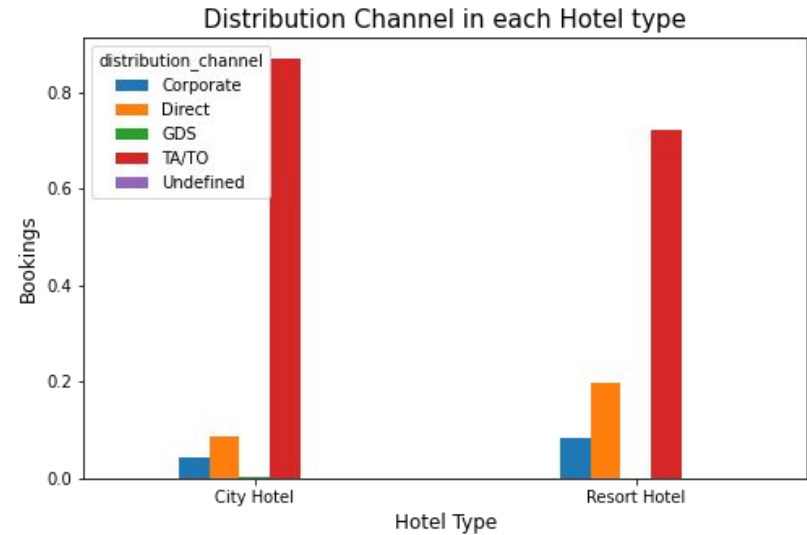
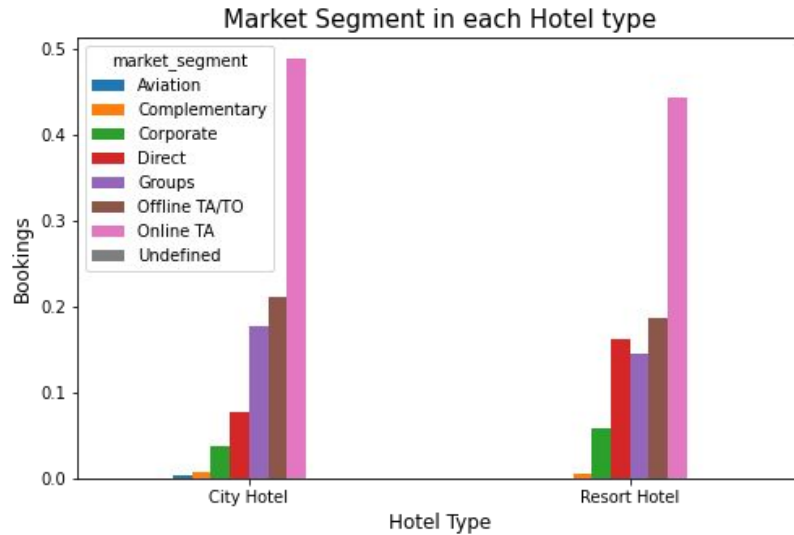


Data Visualization



- City Hotel and Resort Hotel bookings are in the ratio 2 : 1.
- Transient and Transient-Party customers dominated the bookings at both the City Hotel and the Resort Hotel.

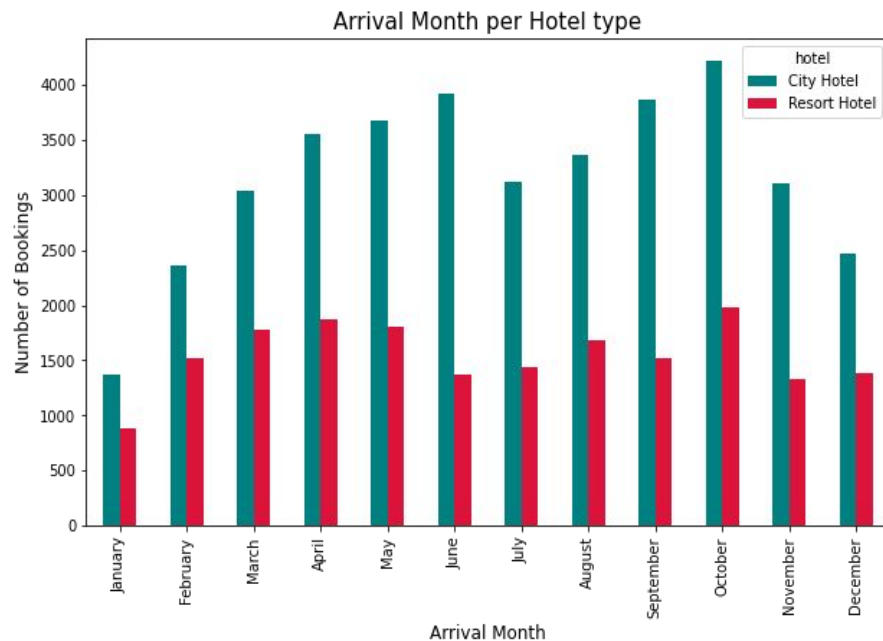
Data Visualization



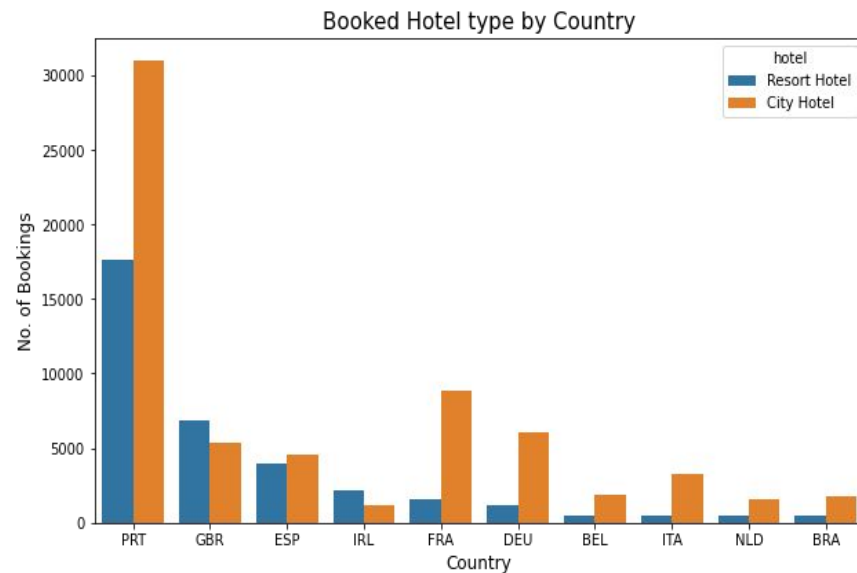
- TA/ TO makes the biggest contribution to bookings, both as a market segment and as a distribution channel.

Data Visualization

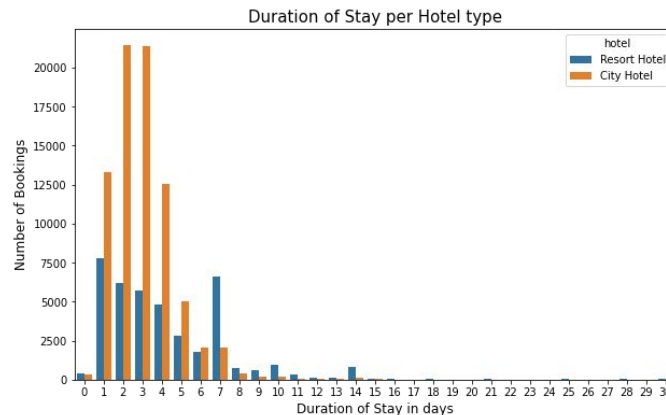
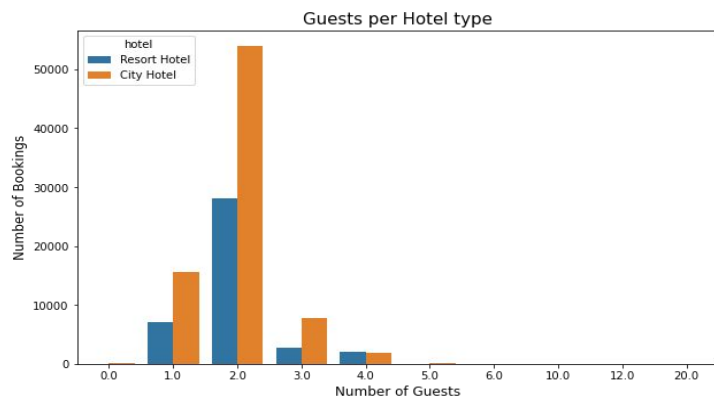
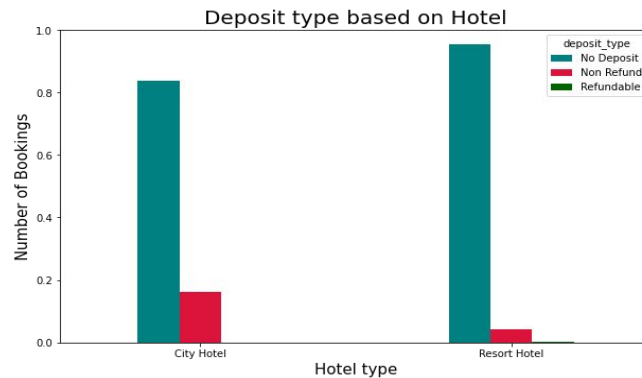
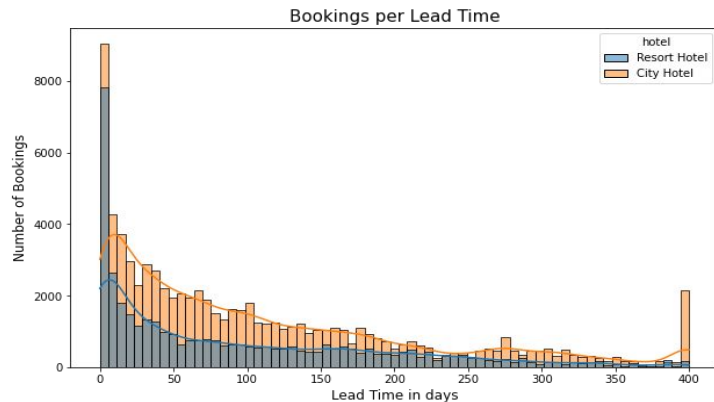
Guests' arrival across the months



Guests' Country of Origin

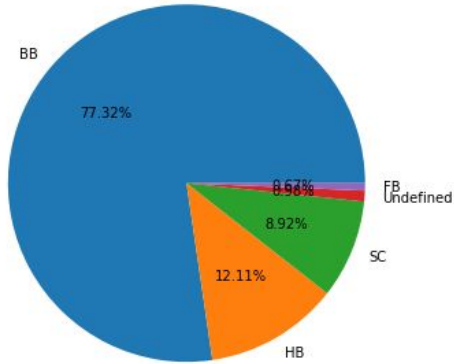


Data Visualization

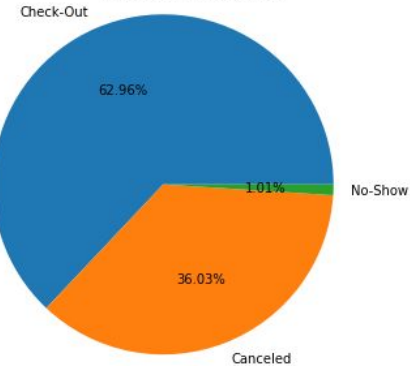


Data Visualization

Bookings with different Meal types



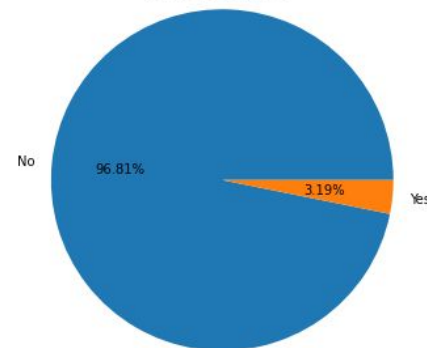
Reservation Status



Average number of Special Requests by Hotel type



Repeat Guest



Exploration of Important Metrics

Booking Cancellations



Customer Type	Hotel Type	Cancellation
Transient	City Hotel	46%
	Resort Hotel	31%
Transient-Party	City Hotel	28%
	Resort Hotel	19.5%
Group	City Hotel	10%
	Resort Hotel	10.5%
Contract	City Hotel	48%
	Resort Hotel	9%

Exploration of Important Metrics

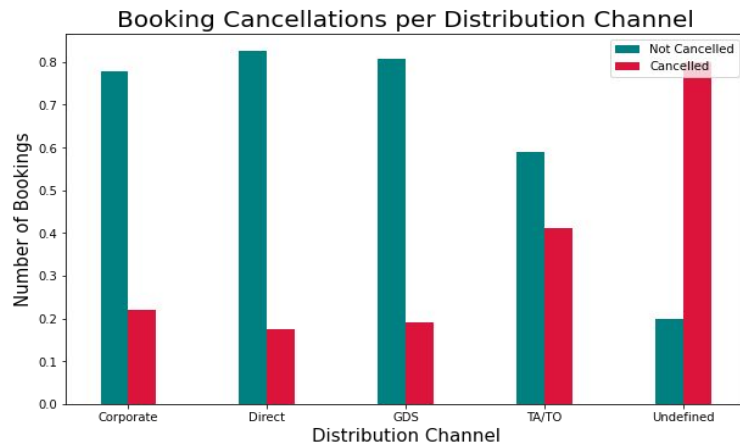
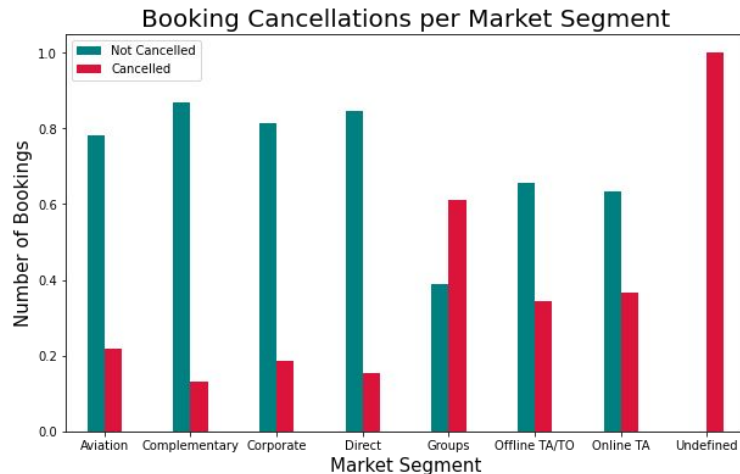
Booking Cancellations

Market Segment

- The Groups segment had highest (61%) cancellations.
- Offline TA/TO and Online TA market segments had 34% and 37% cancellations.
- Direct market segment and Corporate market segment contributed to much lower cancellations at 15% and 18%.

Distribution Channel

- The cancellations were highest (41%) for bookings made through the TA/TO distribution channel.
- Bookings made through the Corporate distribution channel and GDS had 22% and 19% cancellations.
- Direct bookings had the least cancellations at 17%.



Exploration of Important Metrics

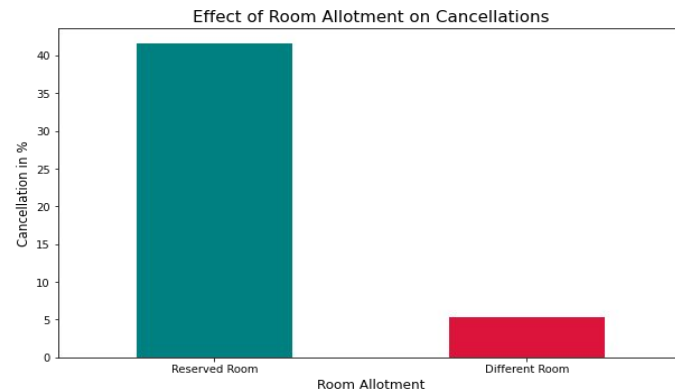
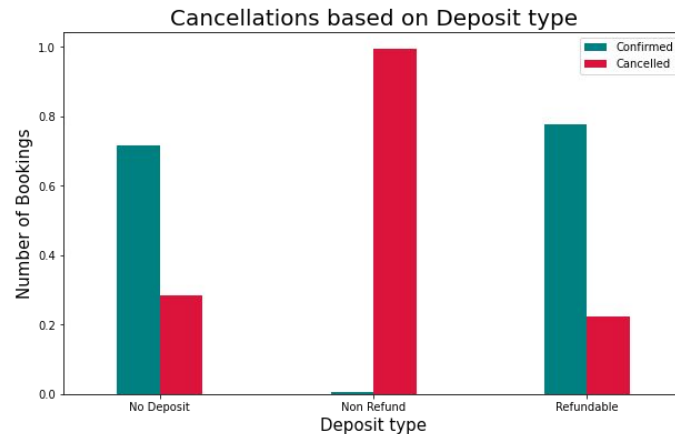
Booking Cancellations

Deposit type

- Bookings made with non-refundable deposits had almost 99% cancellations and those with refundable deposits had 22% cancellations.
- Bookings made without any deposit had 28% cancellations.

Type of Room allotted

- Only 5.4% of the bookings in which a different room was assigned, had cancellations.



Exploration of Important Metrics

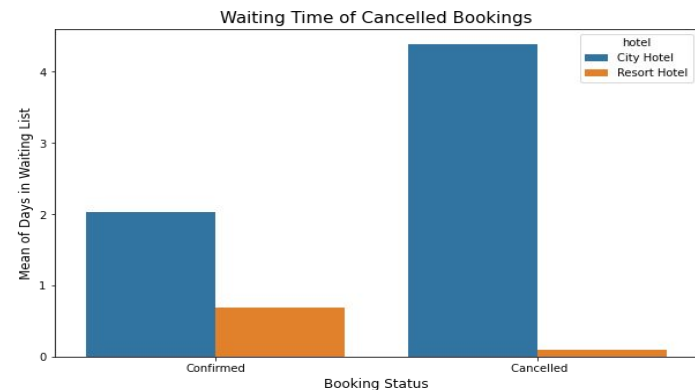
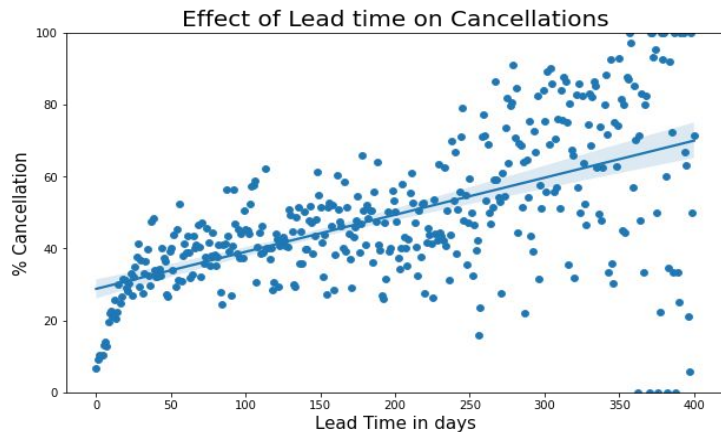
Booking Cancellations

Lead Time

- There is a positive correlation between lead time and booking cancellations. The cancellation % increased as the lead time increased.

Days in Waiting list

- Canceled bookings in City Hotel had a higher mean waiting time.

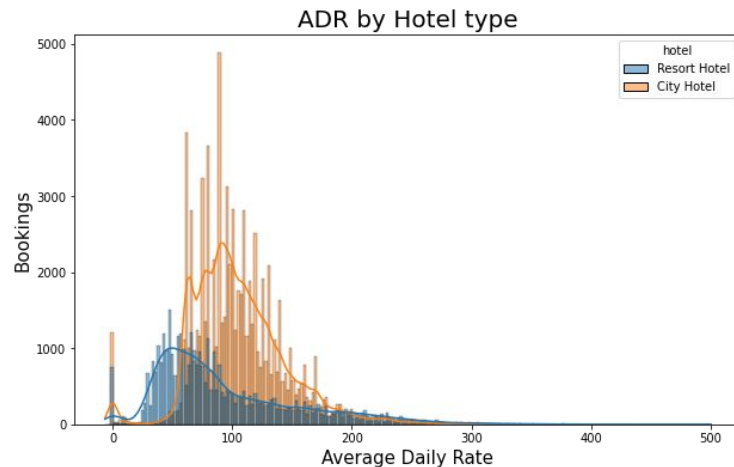
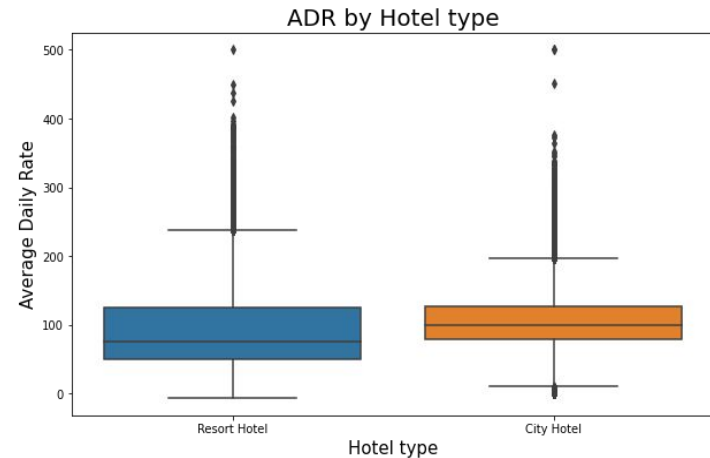


Exploration of Important Metrics

Average Daily Rate (ADR)

ADR is one of the most important metrics in the Hotel industry. It is a measure of the average rental revenue earned for an occupied room per day.

- Mean ADR for City Hotel is 106.73 and for Resort Hotel, it is 96.74.
- The median ADR of City Hotel is 99.96 and Resort Hotel is 78.



Exploration of Important Metrics

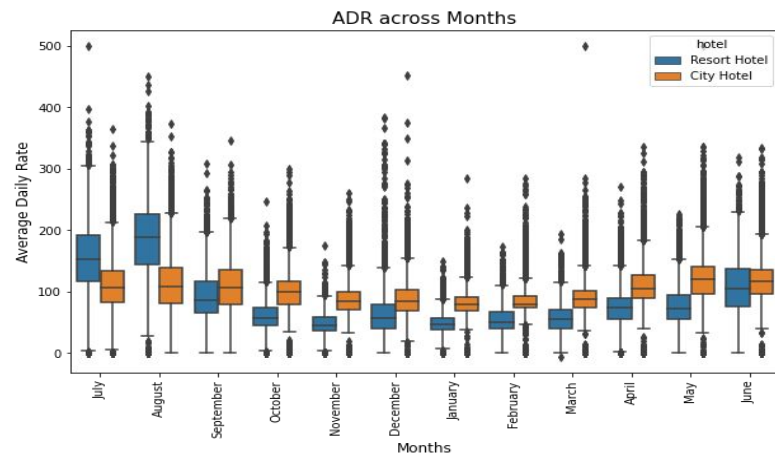
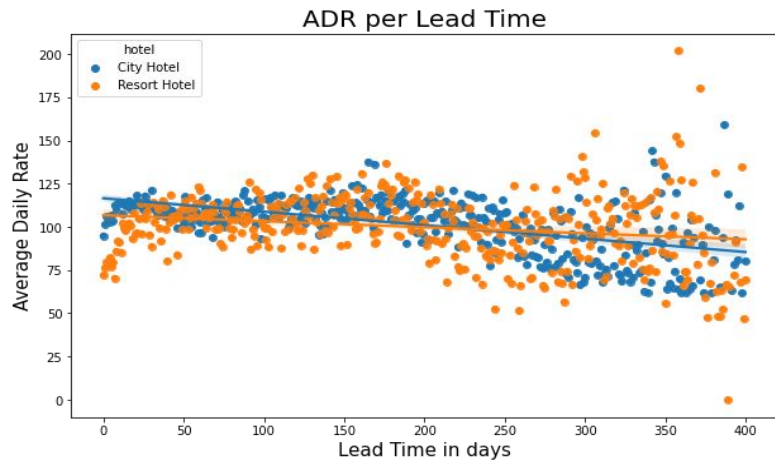
Average Daily Rate (ADR)

Lead Time

- There is a slight downward trend in ADR as the lead time increases.

Months

- The ADR for City Hotel is roughly within the range 80-120 across all the months.
- The ADR for Resort Hotel shoots up in the months of July and August.

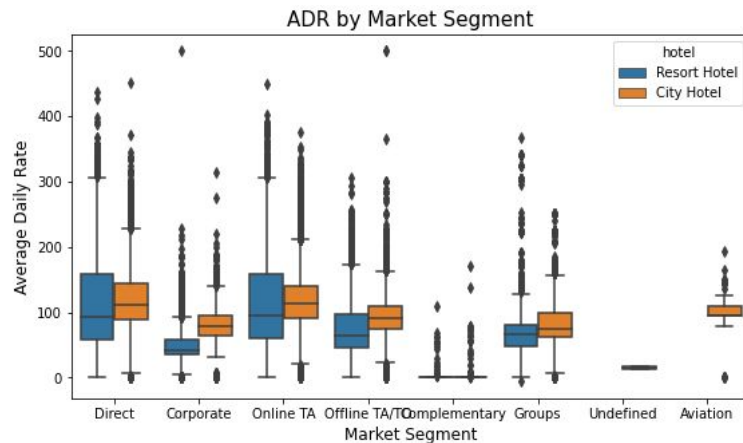


Exploration of Important Metrics

Average Daily Rate (ADR)

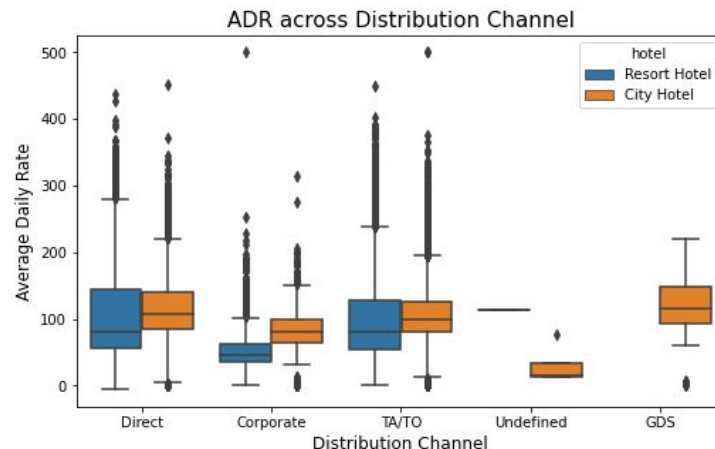
Market Segment

- The contribution to the highest ADR is from the Direct and the Online TA market segments.
- The Corporate market segment brings in the lowest ADR for both types of Hotel.



Distribution Channel

- The ADR for City Hotel is the highest for bookings made through the GDS.
- Direct distribution channel fetches better ADR than TA/TO for both City and Resort hotel.
- Bookings made through the Corporate distribution channel has the lowest ADR.



Exploration of Important Metrics

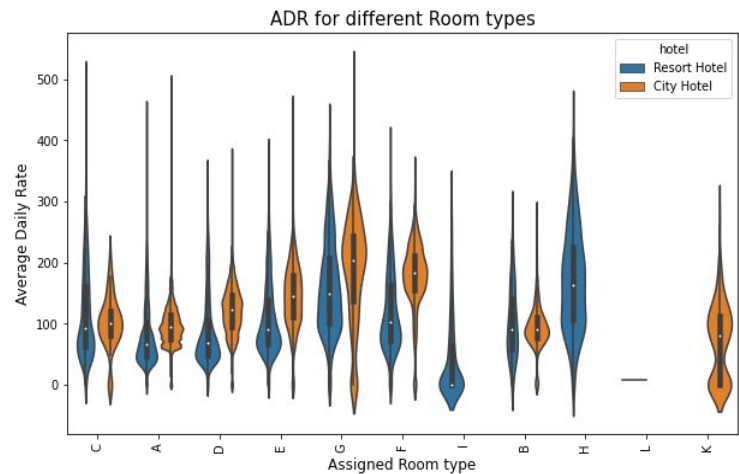
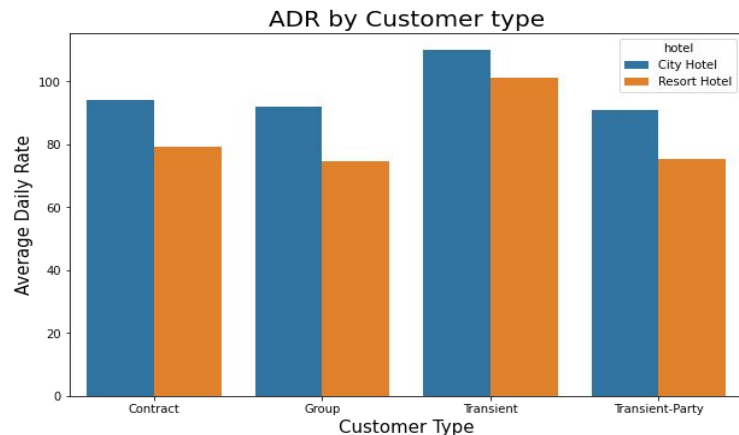
Average Daily Rate (ADR)

Customer Type

- ADR is the highest for Transient customers and lowest for Group customers.

Assigned Room type

- The room types H and G had the highest ADR in Resort Hotel.
- The room types G and F had the highest ADR in City Hotel.



Exploration of Important Metrics

Average Daily Rate (ADR)

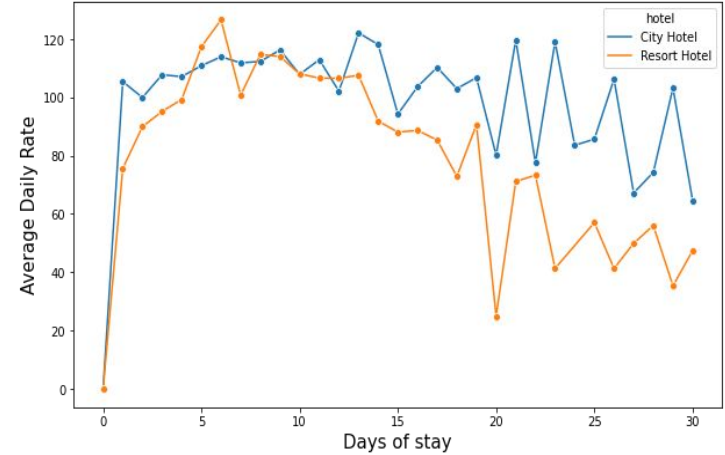
Duration of Stay

- The ADR did not vary much with increase in the duration of stay in City Hotel.
- The ADR for Resort Hotel increased as the number of days increased from 1 to upto 6 days. Beyond 7 days, the ADR begins to fall .

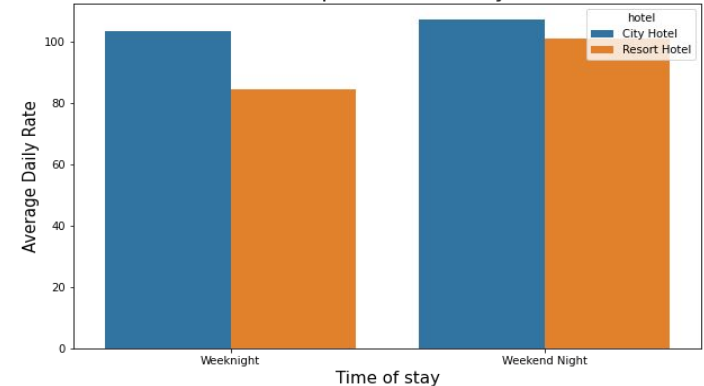
Time of Stay

- The ADR is more or less similar during weeknights and weekend nights for City Hotel.
- ADR of Resort Hotel for stay in weekend nights (100.8) is much higher than that of weeknights (84.3).

ADR per different duration of stay



ADR per Time of Stay



Exploration of Important Metrics

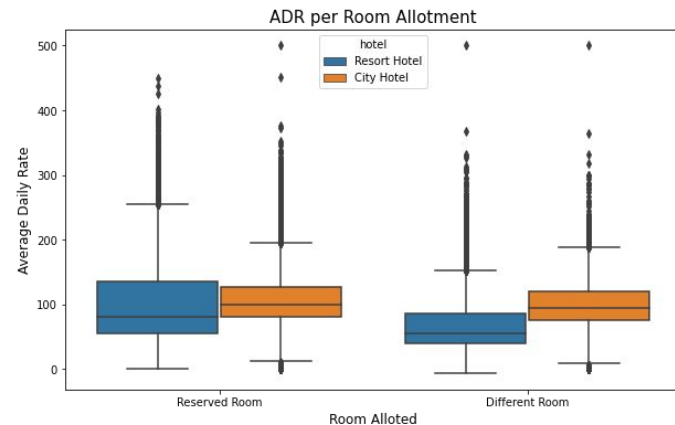
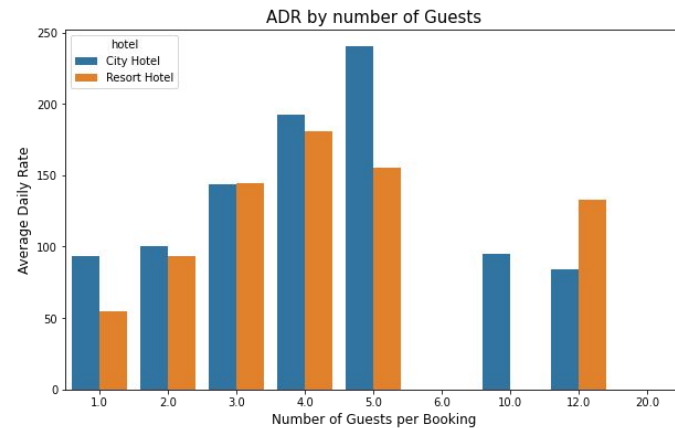
Average Daily Rate (ADR)

Total Guests

- The ADR increases with an increase in the number of guests per booking till at least 5 members.
- For larger groups, the ADR is lower.

Room Allotment

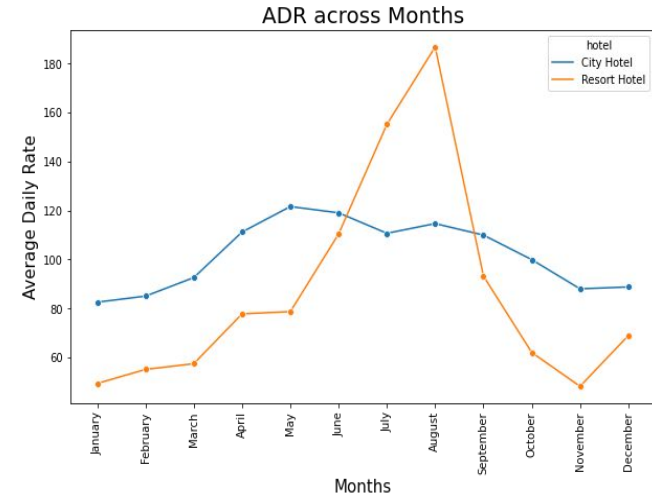
- The ADR is lower if a room that is different from the reserved one, is assigned.



Answering the Questions

When is the best time of the year for the customers to book the Hotel rooms?

- Customers who are price sensitive, may avoid Resort Hotel booking during the months of July and August when the ADR is very high. During the rest of the year, Resort Hotel is much cheaper than the City Hotel.
- Guests can book rooms at a relatively lower price for a stay during the October to April period in City Hotel.



Would room be assigned according to choice?

- Room of choice is assigned to the customer 91% of the time in City Hotel and 81% of the time in Resort Hotel.
- If a different room is assigned, the Resort Hotel offers a better price.

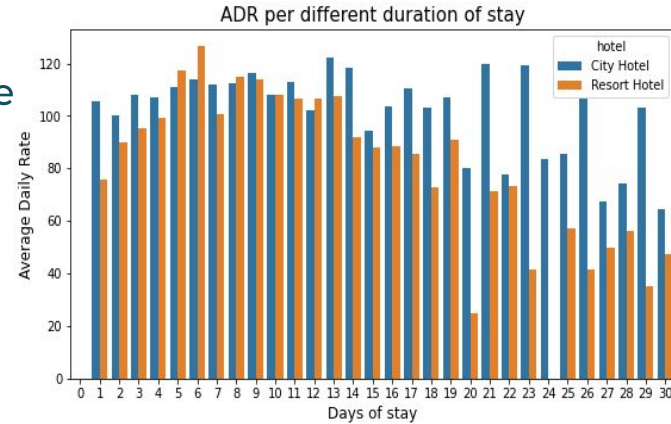
Answering the Questions

What is the optimal length of stay in order to get the best daily rate?

- For Resort Hotel, staying beyond a week would get the customer the best rate. In the City Hotel, daily rate would reduce only for more than 2 weeks of stay.

When is the Hotel more likely to receive high number of special requests?

- The Resort Hotel is likely to receive more special requests during the months of June, July and August. For the City Hotel, it is the months of July, August, November and December when they receive high number of special requests.
- Guests travelling with children, repeat guests, contract customers and group customers tend to make more special requests.



Answering the Questions

How can the ADR be increased?

- Focus on the customer type, market segment and distribution channel that contributes to higher ADR i.e. contract customers, direct market segment, online TA market segment and direct distribution channel.
- Room types H and G at the Resort Hotel, F and G at the City Hotel may be given more importance while assigning rooms.
- Demand for A room type is very high. So, the Hotels may increase their availability.
- Encourage bookings with higher number of guests.
- Promote stays during weekend nights.
- Reduce the cancellations to the maximum extent possible.

Answering the Questions

What are the reasons behind booking cancellations ?

It is difficult to point towards any specific reasons because the overall cancellation percentage is very high, at 42% for City Hotel and 28% for Resort Hotel.

- Lead time has a positive correlation with cancellations. The earlier the bookings are made, the higher the chance of cancellation.
- Days in waiting list may be one of the reasons. Cancelled bookings had a higher mean waiting time in City Hotel.
- Bookings which were assigned rooms that were different from the one reserved, had 5% cancellations.
- Groups as a market segment had higher cancellations. This maybe intuitive because travel plans get jeopardised more often when many members are involved.

Other Insights

- The most reserved (i.e. in high demand) room type is A.
- 19% of the rooms assigned at Resort Hotel are different from the ones reserved by the guests. In the City Hotel, this happened only 9% of the time.
- Resort Hotel has a higher chance that its customer may return for another stay
- Only 12% of the bookings required deposits that were non-refundable. City Hotel required more non-refundable deposits than Resort Hotel.
- Approximately 70% of the bookings in both the Hotels include two guests.
- 85% of the Hotel bookings had no changes.
- 93.8% of the bookings did not require a car parking space.
- The mean waiting time for bookings in City Hotel is 3 days while the wait time for Resort Hotel is less than a day.
- Both, the mean and median lead time is higher for the City Hotel.

Thank You